University of Heidelberg

Department of Economics

Discussion Paper Series | No. 578

Smart or Selfish – When Smart Guys Finish Nice

Johannes Lohse

November 2014

Smart or Selfish - When Smart Guys Finish Nice

Johannes Lohse [a]
University of Heidelberg

November 26, 2014

Abstract

In three different variants of an one-shot public good game I analyze the relationship between cooperation and cognitive abilities, assessed through the cognitive reflection test (CRT). In a between-subjects design, the baseline case is contrasted with two treatment conditions that allow to control for two potentially moderating factors: By employing a test for the presence of confusion, the first condition scrutinizes whether higher cognitive abilities are correlated with cooperation proper or simply grant a better understanding of the incentive structure. The second condition explores the proposition that the link between cognitive abilities and cooperation could depend on the complexity of the decision situation. To exogenously create a cognitively more demanding choice setting, subjects had to decide under time pressure. I find a strong and positive relationship between CRT-scores and cooperation, that is not driven by confusion. Time pressure has a strongly moderating effect on this relationship.

Keywords: Cooperation;Cognitive Abilities;Confusion;Public Goods;Dual Process Theories

[a]Corresponding Author. Email: lohse@eco.uni-heidelberg.de. Postal address: Department of Economics, Bergheimer Str. 20, 69115 Heidelberg, Germany. Phone: +49 6221 548013

1 Introduction

Economic experiments have revealed a high level of unexplained subject-specific behavioral heterogeneity. Given the demographic diversity of participants, there has been growing interest in linking these behavioral differences to differences in observable attributes (e.g. Harrison et al., 2007; Croson and Gneezy, 2009) or personality traits (e.g. Volk et al., 2011, 2012). If such classification was feasible and valid, it could be used to identify segments of the population that, due to their distinct preferences, are most reactive to specific changes in market conditions (Benjamin et al., 2013). In this context the study of cognitive abilities as a measurable attribute with considerable inter-individual variability has recently received increasing attention (e.g. Frederick, 2005; Burks et al., 2009; Oechssler et al., 2009; Benjamin et al., 2013). For example, individuals with higher cognitive abilities have been shown to behave less risk averse (Dohmen et al., 2010), to discount future payments at lower rates (Frederick, 2005; Benjamin et al., 2013), to display smaller biases in financial decision making (Oechssler et al., 2009) and to employ higher levels of strategic reasoning (Burnham et al., 2009; Carpenter et al., 2013).

In comparison to the domains of risk and time preferences, much less is known about the degree to which cognitive abilities can be linked to cooperative behavior. However, a clearer picture of this relationship could be useful in order to understand more about the micro-foundations of the association between cognitive abilities and labor market outcomes (Heckman et al., 2006; Heineck and Anger, 2010). For instance, a positive relationship between cognitive abilities and the propensity to cooperate could explain a fraction of the wage premium received by workers with higher cognitive abilities.

Some of the existing experimental evidence points towards a positive correlation. Burks et al. (2009) find that participants with a higher IQ contribute more in a sequential prisoners dilemma (PD) as first-movers and retaliate more against defection as second-movers. Similarly, Jones (2008) shows that repeated PD experiments run at universities with a higher SAT entry score result in higher average cooperation rates. In the related domain of general social-preferences Chen et al. (2013) establish a positive link between generosity in (non-strategic) dictator and SVO-task experiments and SAT-scores[1]. There is, however, also evidence that goes into the opposite direction. Kanazawa and Fontaine (2013) report higher defection rates among more intelligent participants in a one-shot PD. Furthermore, using the strategy method to uncover different cooperative types in a public good game (PGG), Nielsen et al. (2014) find that strict free-riders have higher scores in the cognitive reflection test (CRT) (Frederick, 2005) than conditional or unconditional cooperators.

This paper complements the existing literature by being the first to scrutinize the relationship between cooperation in a one-shot PGG and cognitive abilities measured by the CRT. Furthermore, in light of the as yet inconclusive evidence, it employs two randomly assigned treatment conditions to identify if the proposed link holds, irrespective of two potentially moderating factors that have been left unexplored so far.

The first treatment condition addresses the concern that some subjects in public good games could be confused about the true incentive structure (Andreoni, 1995; Houser and Kurzban, 2002; Ferraro and Vossler, 2010; Bayer et al., 2013). For such confused subjects contributions cannot simply be equated with cooperation. Moreover, confusion is plausibly in itself a function of cognitive abilities. Therefore, in order to distinguish between both contribution motives, the baseline condition is contrasted with a public good variant (Houser and Kurzban, 2002) in which preferences for cooperation cannot drive behavior.

Furthermore, there is reason to believe that a link between deliberative capacities - as assessed by the CRT - and cooperative behavior could vary with the complexity of the decision situation (Duffy and Smith, 2014). The second treatment condition directly tests this proposition, by exogenously making the decision situation cognitively more

[1] For a deviating result for female participants see Ben-Ner et al. (2004)

demanding through reducing the time available for deliberation. Thereby I look into the question, whether cognitive abilities correlate with social preferences independently from the decision environment or whether an increase in cognitive complexity has a moderating effect.

In line with the first strand of literature I find a significantly positive correlation between CRT-scores and contribution behavior in the baseline PGG. According to the first treatment condition, this result is not driven by varying levels of confusion for subjects with distinct CRT-scores. On the other hand, applying time-pressure has a strongly moderating effect on the significance of the correlation: For subjects in the second treatment condition there are no detectable differences of contributions related to their cognitive abilities.

The remainder of this paper is structured in the following way: In section 2 I describe the experimental design. Section 3 contains results and robustness checks. Section 4 closes with a short discussion of the findings.

2 Methods and Procedures

2.1 Measuring Cognitive Abilities

As a measure of cognitive abilities I used the cognitive reflection test (CRT) as introduced in Frederick (2005). This three-item test assesses subjects' predisposition to base their decisions on intuitive rather than deliberative thinking. To discriminate between these cognitive styles, each of the three test items is a question that has an intuitive but incorrect answer, while the correct answer can be easily found with a sufficient amount of reflection. Despite this simple setup, CRT-scores have been shown to be significantly correlated with broader measures of cognitive abilities such as scores from the *SAT*, the *Wonderlic Personnel Test* or the *Wechsler Matrix Test* (Frederick, 2005; Toplak et al., 2011, 2014). To increase the reliability of measured scores and overcome the potential issue that the intrinsic motivation to reach a good test result could vary across subjects (Chen et al., 2013), participants could earn a monetary reward of up to four Euros for giving a correct answer to each question. There are some concerns (Toplak et al., 2014) that by now subjects could be familiar with some of the CRT items due to the test's rising popularity in research. To address these concerns only subjects with limited prior lab experience[2] were invited to take part in the experiment and I collected information on how familiar the participants were with the CRT. This information is used in a robustness check.

2.2 Measuring Cooperation

In total I employed three different variants of a public good game to analyze the link between cooperation and cognitive abilities. Subjects were randomly assigned to one of these three conditions in a between-subjects design. Each of the two treatment conditions is contrasted against a baseline condition. This comparison is used to explore the role of two moderating factors that could potentially influence the link between cognitive abilities and cooperation.

Baseline

The baseline task was a standard linear public goods game as reviewed in Ledyard (1994) and Zelmer (2003), based on the common payoff formula[3] ($\pi_i = 0.2(\omega - x_i) + 0.1x_i + 0.1\sum_{j\neq i}^{N=3} x_j$): Subjects were randomly and anonymously matched into groups of four and received an initial endowment of 20 tokens each worth 0.20 Euro in the private account.

[2]The median subject had participated in only two prior studies

[3]π_i denotes the payoff to subject i, ω is the initial endowment and x_i is the contribution of subject i.

Contributions to the public account created a payoff of 0.10 Euro for each subject in the group resulting in a MPCR of 0.5. Given these incentives, free-riding is a dominant strategy while full contributions would maximize the total payoff for the group. The game was one-shot and there was no feedback given to other group members.

Variant 1: Disentangling Different Contribution Motives

There are justified concerns that contributions observed in a public good game cannot be directly equated with cooperative behavior, as a significant number of subjects contributes out of confusion. Based on different methods several studies (Andreoni, 1995; Houser and Kurzban, 2002; Ferraro and Vossler, 2010; Bayer et al., 2013) have established that some subjects lack a basic understanding of how their choices, together with the choices of other group members, affect experimental payoffs for themselves and for their group. Hence, in absence of an appropriate control condition it would be unclear, how to interpret a link between cognitive abilities and contributions observed in the baseline. For instance, a negative correlation could be seen as evidence that subjects with a higher CRT-score are genuinely less cooperative or conversely that they are simply more able to understand the underlying incentive structure and hence choose their dominant strategy of defection. To differentiate between these interpretations, I used a variant of the public good game initially introduced by Houser and Kurzban (2002) to study the extent of confusion in public good games. This variant retains all features of the baseline condition with the only difference that the gains from cooperating are removed: Subjects shared their public account with a computer program instead of sharing it with human interaction partners. With an MPCR of 0.5, subjects lost money for each token contributed without generating additional gains for other experimental subjects. Thus contributions in this variant cannot be attributed to cooperative motives vis-á-vis these subjects. To successfully exclude other reasons for contributing in variant 1 it is essential that subjects understand the difference between human and computerized interaction partners and acknowledge that the computer is not programmed to react to their amounts chosen. To ensure this, they were informed that the computer agents would contribute a predetermined amount and (naturally as they are only a computer program) would not receive any payoffs generated from the public account. The predetermined contribution by the computer agents had been written on a concealed poster in the room prior to the experiment and was revealed to subjects at the end of the session. Subjects were informed about this procedure before making any decision in this task. A manipulation check confirmed that 92 percent of subjects understood that they had interacted with a computer program and 93 percent believed that they were not able to influence the computer's contribution.

Variant 2: Increasing the cognitive demand

Choice situations offering an opportunity for cooperation vary by the level of cognitive demands imposed on the decision maker. On the one hand, cognitive demands could increase when the situation in itself becomes more complex and hence cooperative choices are harder to recognize. For instance, in an experimental context moving from a two player PD to a n-player linear PGG could make the choice situation cognitively more demanding, as the strategy space gets larger. On the other hand, cognitive demands could be higher in situations in which it is more difficult for the decision maker to make full use of his cognitive abilities. Such situations could arise if the decision maker is engaged in multiple parallel tasks or has to come to a conclusion under time pressure. Clearly, the link between cognitive abilities and the level of cooperation could be affected in both cases (Duffy and Smith, 2014). This study only investigates the second case, as increasing the complexity of the cooperation task through enlarging the strategy space - while interesting in principle - would in most meaningful ways complicate the comparison to the baseline PGG.

In variant 2, as a way to make the choice situation cognitively more demanding, subjects were therefore asked to decide within a short time limit, while keeping the basic PGG task itself unchanged. Under time pressure the use of deliberative resources is more demanding and thus decisions based on intuition are more likely to occur (Rand et al., 2012, 2014). While there is mixed evidence[4] on the direct effects of time pressure or cognitive load in PGG or PD games, it is also unclear if these manipulations affect participants with different cognitive abilities in a uniform way.

Subjects under time pressure were asked to decide within 7 seconds, a limit constructed by subtracting one standard deviation from mean decision time in the first two baseline sessions. In order to keep them from adapting to the time pressure by already choosing a contribution level while reading the instructions, they were only informed about the time pressure right before seeing the decision screen. A counter on the screen showed the remaining time.

2.3 General Procedures

The experiment was conducted at the University of Heidelberg Experimental Computer Lab between December 2012 and March 2013 during 21 sessions. The 284 participants were recruited from a standard subject pool of undergraduate and graduate students using ORSEE. The subjects were from mixed disciplines, including economics. Subjects that had previously taken part in a public goods experiment were excluded from recruitment to the experiment. Upon arrival, participants were seated at their computer terminal, generated a random password to ensure their anonymity and received a set of general instructions that were read aloud by the experimenter. All other instructions, tasks and questionnaires were fully computerized using z-Tree. For all subjects the CRT was conducted after the public goods task and several questionnaires. At the end of the experiment, participants were paid their earnings in private. All sessions lasted approximately 75 minutes and participants earned an average of 9,55 Euro (Min.:5 Euro;Max 15.00 Euro), including a show-up fee of 3 Euro[5].

3 Results

3.1 Observed Behavior

Table 1 displays the main results of the cognitive reflection test. As expected under random assignment, there are no large differences between the different treatment conditions. Comparing the single items shows that the first item (Ball and Bat) is answered correctly less often than the other two items. The last row shows an overall CRT-score that is constructed by summing up the number of correct answers given by each subject. On

Table 1: CRT-Score and Single Items across Treatments

	Baseline (N=108)	Variant 1 (Confusion) (N=64)	Variant 2 (Cognitive Demand) (N=112)	Overall (N=284)
Item 1 (Ball and Bat)	0.46 (0.40)	0.43 (0.37)	0.59 (0.53)	0.51 (0.42)
Item 2 (Widget)	0.62 (0.59)	0.67 (0.62)	0.60 (0.55)	0.62 (0.55)
Item 3 (Lilly Pad)	0.70 (0.63)	0.66 (0.64)	0.68 (0.61)	0.68 (0.61)
Total CRT-Score	1.78 (1.54)	1.76 (1.64)	1.87 (1.6)	1.81 (1.58)

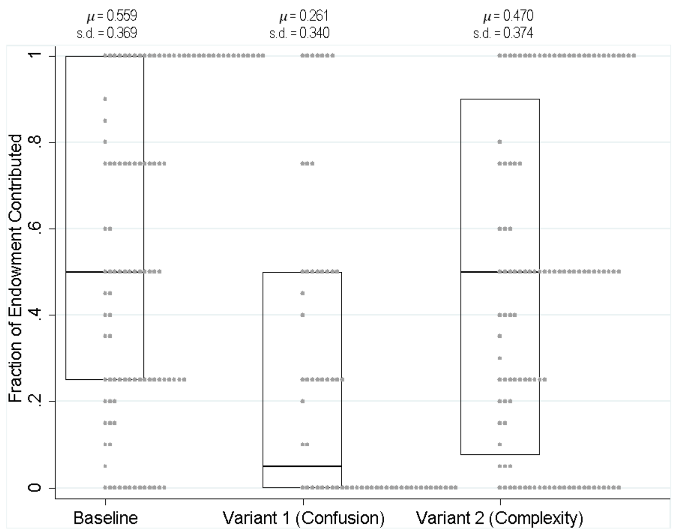
Notes: CRT-scores across the different treatment conditions. Values in the brackets display scores when discarding observations from subjects having indicated to know this specific CRT item.

[4]While Rand et al. (2012, 2014) find a significant increase of one-shot cooperation under time pressure, this result is not replicated by Tinghög et al. (2013). Duffy and Smith (2014) even find weak evidence for less cooperation under cognitive load in a repeated setting.

[5]There were also other incentivised questionnaires and decision tasks not analyzed in this paper

average 1.81 of the questions are answered correctly, which is within the upper range of values reported in Frederick (2005). This average results from the following distribution of individual CRT-scores: 34.9 percent of subjects answer all items correctly, 28.5 percent of subjects answer two items correctly, 20.1 percent of subjects answer one item correctly and the remaining 16.5 percent of subjects answer none correctly. The values in the brackets show solution rates when discarding observations from those 12-19 percent of subjects who indicate in the follow-up questionnaire that they already knew the solution of a given item.

Figure 1: Box plots of contributions across PGG variants



Notes: This graph shows the fraction of endowment contributed separately for the three PGG variants. The black line indicates median contributions. The lower and upper quartiles are marked by the surrounding box. The stacked gray dots can be read as a histogram of individual contributions plotted in a horizontal direction.

Figure 1 displays the distribution of contributions across the three different PGG variants. As commonly observed in the literature (Ledyard, 1994), for the baseline average contributions fall into the range of 40 to 60 percent of initial endowment. Furthermore, there are three spikes in the distribution indicating that more than half of the participants either free-ride, contribute their full endowment or choose to split it between the private and public account.

For variant 2 the distribution of contributions is close to the baseline with an identical median, and a higher incidence of free-riding. In contrast to Rand et al. (2012, 2014) but in line with Tinghög et al. (2013) the application of time pressure results in a contribution average below baseline at a weakly significant level (M.W. rank sum p=0.0968).

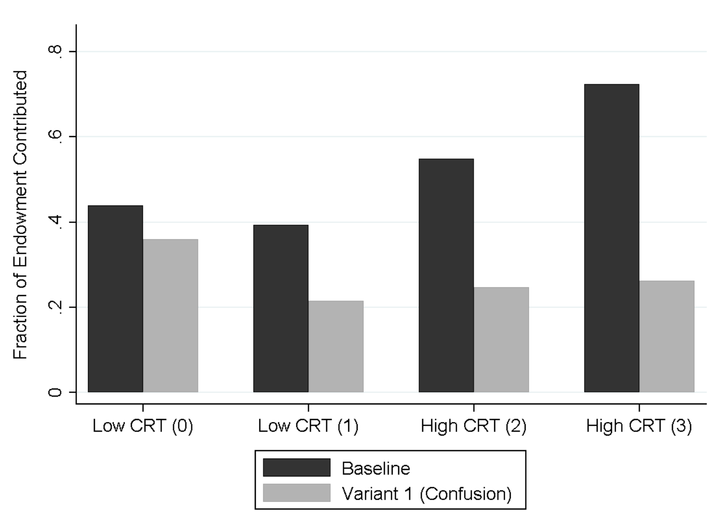
In variant 1 behavior deviates substantially from the baseline. When interacting with a computer program, subjects on average contribute 0.261 percent of their endowment, well in line with values observed in Houser and Kurzban (2002). This reduction, relative to the baseline, is mainly driven by the 50 percent of subjects who show no sign of confusion and contribute zero tokens. The positive contributions of the other half of subjects can be interpreted as indicating the presence of confusion (Houser and Kurzban, 2002).

3.2 Comparing Baseline and Variant 1

Figure 2 displays the relationship between the number of correct answers in the CRT and contributions to the public account, separately for the baseline (black bars) and variant 1 (gray bars). Contributions in the baseline public good game can be interpreted as an expression of cooperative preferences but also, in part, as being the result of confusion about the underlying incentives. In contrast, in variant 1 in which subjects' contributions go to a public account shared with a computer program, cooperative preferences towards other group members cannot motivate behavior. A comparison of both conditions can thus be used to differentiate between a link between CRT-scores and cooperative preferences and a link between CRT-scores and confusion.

For the baseline condition there is a positive relationship between CRT-scores and contributions, which is highly significant (Spearman's Rho = 0.3383, p= 0.0003). On average, subjects with the lowest possible CRT-score contribute 44 percent of their initial endowment to the public account while subjects with the highest CRT-score contribute 73 percent. This large increase by two-thirds is driven by the growing fraction of subjects who contribute their full endowment, while the fraction of free-riders remains close to constant across different CRT-score categories. Only 15 percent of subjects in the lowest CRT-score category contribute their full endowment, while this is true for 50 percent of subjects in the highest CRT-score category. Comparing contribution behavior across all CRT categories furthermore shows that there are significant differences in average behavior for subjects grouped into different CRT-score categories (ANOVA: F=5.17, p=0.0023).

Figure 2: Average Contributions by CRT-Scores: Baseline vs. Variant 1



Notes: This graph shows the fraction of endowment contributed grouping subjects by their CRT score. Black bars are used for baseline behavior and gray bars for behavior in variant 1.

In variant 1 of the PGG there is a negative yet insignificant (Spearman's Rho = -0.1289, p= 0.3099) correlation between contributions and CRT-scores: On average, subjects within the lowest CRT category contribute 36 percent of their initial endowment and this average drops to approximately 25 percent for subjects in the other three CRT categories. When comparing all CRT-score categories there is no evidence for significant differences in average contributions. (ANOVA: F=0.38, p= 0.7711).

If contributions in variant 1 are mainly driven by subjects' confusion status as proposed in the literature (Houser and Kurzban, 2002; Ferraro and Vossler, 2010)[6], then

[6]Other candidate explanations include experimenter demand effects, a misunderstanding of the rules

especially low CRT subjects display confused behavior. Within this group only 30 percent of subjects contribute zero tokens as compared to 57 percent of subjects in the highest CRT-score category.[7] Given that the link between CRT-scores and contributions goes in opposite directions in the baseline and variant 1, the positive relationship observed in the baseline is implausibly driven by a link between confusion and cognitive abilities. Rather confusion could mask an even stronger relationship between cooperation and CRT-scores. Baseline contributions for subjects with the lowest CRT-score are statistically indistinguishable (M.W. rank sum p=0.4179) from contributions in variant 1. Hence, relative to subjects in higher CRT categories only a smaller fraction of baseline contributions of low CRT subjects can be attributed to cooperative motives and a larger fraction to confusion. In contrast, for the two highest CRT-score groups baseline contributions are significantly larger (M.W. rank sum $p < 0.01$). This change of the relative importance of contribution motives across different CRT groups could bias the baseline correlation between contributions and CRT-scores downward, so that it can serve only as a lower bound for the true relationship between cooperation and CRT-scores.

One additional concern is that high CRT-scores could overstate the true cognitive abilities for those subjects already familiar with parts of the test (Toplak et al., 2011, 2014). As a robustness check, I thus repeat the previous steps of analysis but restrict attention to the subset of subjects who are not familiar with more than one test item[8]. This procedure does not affect the main conclusions. For the baseline there is still a high and significant correlation (Spearman's Rho = 0.3308, p= 0.0011) and average contributions are significantly different between all CRT categories (ANOVA: F=4.73, p=0.0041). For variant 1, in contrast, I find no significant correlation (Spearman's Rho = -0.1003, p= 0.4706) and subjects do not show different average contribution levels according to their CRT score (ANOVA: F=0.80, p=0.4990).

Taken together, this evidence suggests a significantly positive relationship between cognitive abilities and cooperative behavior, while ruling out confounds due to ignoring confusion as an alternative interpretation.

3.3 Comparing Baseline and Variant 2

Figure 3 plots the relationship between the number of correct answers in the CRT and contributions to the public account. Average contributions are displayed as black bars for the baseline and gray bars for variant 2. By comparing these conditions I analyze whether the link between cognitive abilities and contribution behavior is moderated by the cognitive demand imposed by a given choice situation. The CRT assigns a high score to a decision maker, who easily overrides a choice resulting from intuition in favor of a choice that is based on a more deliberative assessment of the different options. Conceivably, under time pressure, subjects with a high CRT-score can make less use of their higher deliberative capacities as employing deliberation requires more time relative to intuition (Evans, 2008). Therefore, I expect that differences in CRT-scores can explain a smaller fraction of the variation in contribution behavior in variant 2 than in the baseline.

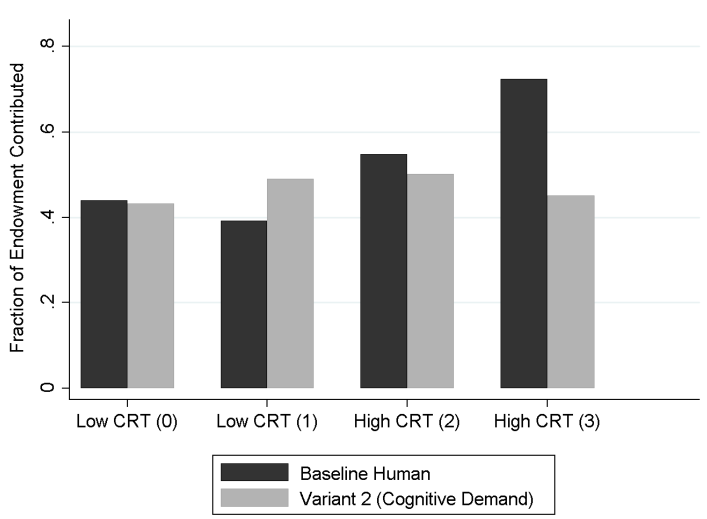
As analyzed in detail in the previous section, there is a strong positive relationship

by which the computer program acts or altruism towards the experimenter. While the latter two factors do not seem to be of particular (ex-post) relevance to the subjects, given their answers in a follow-up questionnaire, there is no additional information on the relevance of the first factor. Mark, however, that unless one assumes that experimenter demand effects are stronger in one of the conditions, this does not affect the comparison of both conditions.

[7] Another observation that plausibly supports the notion of higher confusion levels among subjects in the lowest CRT category, is that for this group the proportion of choices that do not fall in one of the three distributional spikes (at 0 percent, 50 percent or 100 percent of endowment) is substantially larger than in the highest CRT group. This is true both for the baseline and variant 1 (Baseline: 70 perc. vs. 26 perc.; Variant 1: 50 perc. vs. 19 perc.). If confused subjects are more likely to select a random contribution from the available range, this large difference could be seen as an additional indication of confused behavior.

[8] Based on their answers in a follow-up questionnaire.

Figure 3: Average Contributions by CRT-Scores: Baseline vs. Variant 2



Notes: This graph shows the fraction of endowment contributed grouping subjects by their CRT score. Black bars are used for baseline behavior and gray bars for behavior in variant 2.

between CRT-scores and contributions in the baseline condition. On average, subjects with the highest CRT-score contribute 66 percent more than subjects with the lowest CRT-score.

Turning to behavior in variant 2 by contrast, there is no significant correlation (Spearman's Rho = -0.0221, p = 0.8171) between CRT-scores and contributions. Furthermore the fraction of subjects contributing their full endowment is comparable for the highest (27 percent) and lowest (12 percent) CRT-score category (M.W. rank sum p = 0.1992) and average contributions do not differ between all CRT-score groups (ANOVA: F = 0.19, P=0.9050). The incidence of free-riding is also not linked to CRT- scores.

Comparing behavior between variant 2 and the baseline indeed confirms that subjects with higher cognitive abilities - as measured by the CRT - do not behave differently, when the decision environment constrains the use of these abilities: In comparison to the baseline, contributions under time pressure are at significantly lower levels for subjects with high CRT scores of 2 or 3 (N = 142; M.W. rank sum p = 0.0101). At lower CRT scores of 0 or 1, however, subjects in variant 2 did not contribute different amounts compared to those in the baseline (N = 78; M.W. rank sum p = 0.4574). Excluding subjects with prior knowledge of the CRT does only marginally affect these results.[9]

Taken together these findings demonstrate that subjects with higher deliberative capacities are more cooperative, if the decision environment allows for the use of these capacities. If this is not the case contributions across all CRT groups converge to the level of low CRT subjects. This overlooked interaction effect could explain some of the conflicting findings in the literature on the effects of time pressure in public good games (Rand et al., 2012, 2014; Tinghög et al., 2013), if the single studies are based on samples with strongly diverging CRT-score distributions[10].

[9] For high CRT subjects differences remain weakly significant (N = 119; M.W. rank sum p = 0.0808). For only the highest CRT category differences stay significant (N = 62; M.W. rank sum p = 0.0493)

[10] Frederick (2005) provides some evidence how CRT-score distributions can already differ significantly between student subjects from different academic institutions

3.4 Regression Results

It is well documented that CRT-scores are significantly related to a number of other individual attributes such as risk preferences, general cognitive abilities or gender (Frederick, 2005; Oechssler et al., 2009). Subjects in this study are no exception: Male participants reach a significantly (M.W. rank sum p = 0.0029) higher average CRT-score (2.01) than female participants (1.63). Surprisingly high and low score subjects do not differ in their (stated) risk attitudes. This could be due to assessing risk preferences by an unincentivised survey question, even though this questionnaire item has been shown to have a high behavioral validity (Dohmen et al., 2011). Furthermore, CRT-scores are significantly correlated with a proxy of general cognitive abilities, namely better grades[11] in the math and language section of German A-level exams (Spearman's Rho = -0.2709, p = 0.0000). To the degree that each of these attributes is also related to contribution behavior in public good games (Croson and Gneezy, 2009; Teyssier, 2012) the previously reported correlations could be biased. By employing a multivariate regression framework I thus estimate the partial correlation of higher CRT-scores with contributions, holding constant a set of observable characteristics. Obviously, this method cannot exclude the possibility that CRT-scores are related to an unobserved characteristic driving the results.

Table 2: Regression Results

	(1) Baseline		(2) Variant 1		(3) Variant 2		(4) Full Sample		(5) Unconfused	
CRT-score (0-3)	0.0854**	(2.29)	-0.0289	(-0.63)	0.0186	(0.50)	0.110***	(3.22)	0.143***	(3.17)
Male (1=Yes)	0.198**	(2.25)	0.0681	(0.62)	0.0154	(0.20)	0.0763	(1.54)	0.0772	(1.17)
Risk (1-11)	0.0157	(1.02)	0.00134	(0.07)	-0.0114	(-0.72)	0.00200	(0.21)	0.0131	(1.03)
Average Grade (1-6)	-0.0312	(-0.70)	-0.0163	(-0.25)	-0.0530	(-1.03)	-0.0355	(-1.18)	-0.0720*	(-1.87)
Age (Years)	0.00604	(0.44)	-0.0117	(-0.82)	-0.00857	(-0.50)	-0.00638	(-0.69)	-0.00607	(-0.45)
Master Student (1=Yes)	-0.0737	(-0.63)	-0.193*	(-2.00)	0.0435	(0.51)	-0.0480	(-0.82)	-0.126*	(-1.70)
Economics Major (1=Yes)	-0.00450	(-0.05)	0.0746	(0.67)	0.0360	(0.40)	0.0380	(0.69)	0.0805	(1.04)
Prior Lab Experience (1=Yes)	0.00568	(0.08)	0.163	(1.63)	-0.0653	(-0.69)	0.0150	(0.30)	0.0185	(0.25)
Known CRT-items (0-3)	0.0635	(1.52)	0.0315	(0.56)	-0.0789*	(-1.93)	-0.00576	(-0.21)	0.0177	(0.47)
Variant1(1=Yes)							-0.0570	(-0.55)	0.129	(0.98)
Variant1*CRT-score							-0.139***	(-2.62)	-0.207***	(-3.22)
Variant2(1=Yes)							0.122	(1.25)	0.196	(1.52)
Variant2*CRT-score							-0.114**	(-2.49)	-0.161**	(-2.55)
Constant	0.243	(0.84)	0.730*	(1.76)	0.794**	(2.10)	0.592***	(2.82)	0.609**	(2.08)
Observations	83		55		102		240		136	
R2	0.2514		0.1903		0.0538		0.1700		0.2349	
Prob > F	0.0001		0.2148		0.6712		0.0000		0.0000	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Notes: This table shows OLS coefficients for the CRT-score when controlling for a set of individual attributes. T-statistics based on robust standard errors are shown in parentheses. The lower number of observations results from incomplete observations especially for the variable *average grade*. Excluding this variable does not change results. Using a Tobit Model as an alternative specification that accounts for potential censoring in the contribution data does also not alter the results for CRT-scores.

Table 2 shows several specification of an OLS regression with contributions as a fraction of the initial endowment being the dependent variable. Specifications (1) - (3) estimate separate models for the baseline and each of the treatment conditions. In addition to the potentially confounding factors I describe above I also control for several variables that could have an impact on how well subjects understand the PGG task. Holding constant these variables does not change the conclusions from the previous sections. CRT-scores are significantly correlated with contributions in the baseline, but in neither of the treatment conditions. With few exceptions none of the control variables is related to behavior in the different PGGs either. Male subjects contribute higher amounts in the baseline and master's students give less when interacting with computer agents, which could reflect a better understanding of the incentives. Specification (4) uses the full sample and tests for heterogeneous effects of cognitive abilities by including treatment dummies for

[11] According to the German convention the best grade is coded as a low number (1) and the worst grade as a high number (6). Consequently, a negative correlation indicates that higher CRT-scores are related to better grades.

assignment to variant 1 or variant 2 and their interaction with CRT-scores. Again, previous results persist. The main effect of CRT-scores remains positive and significant, while each of the interaction effects is significantly negative. Thus, compared to the baseline, the relationship between CRT-scores and contributions is significantly smaller and even reverses in sign for variant 1. In section 3.2 I propose that the true correlation between baseline contributions and CRT-scores could be even larger if all subjects were unconfused. This statement relies on a between-subjects comparison between the baseline and variant 1. Specification (5) provides a robustness check using a different method. After finishing their contribution decision each participant was asked to identify the strategy that they believed would have maximized their own payoff, independent from their actual choice. Based on this incentivised question subjects can be categorized as unconfused if answering zero and confused otherwise. Overall, only 55 percent of subjects were able to name their dominant strategy as the correct answer. Thus, in specification (5) only subjects categorized as unconfused are included in the regression. This does not affect the relationship between CRT-scores and contributions. Rather, as proposed, the relationship gets slightly stronger. Interestingly, when excluding confused subjects, the relationship between grades and contributions becomes also weakly significant showing that subjects with better (i.e. lower) grades tend to contribute more. This replicates findings from studies based on SAT scores[12] and less complex sharing tasks, where confusion arguably plays a smaller role (Jones, 2008; Chen et al., 2013).

4 Discussion and Conclusion

This study provides further evidence that cognitive abilities are an important factor when systematically structuring the heterogeneity of preferences observed in economic experiments. I find a positive correlation between cognitive abilities and cooperation in a PGG. As for all correlational evidence, this finding should be interpreted with some level of care, as it cannot be ruled out that unobserved factors drive parts of the reported relationship, even when experimentally controlling for two important moderating factors. The optimal empiric strategy to get around this potential bias and establish causality would be to randomly assign cognitive abilities to different individuals, which is infeasible for obvious reasons. Instead, like in the PGG (Variant 2) of this study, cognitive load (Duffy and Smith, 2014) or time pressure (Rand et al., 2014) have been used as methods to inhibit the use of existing cognitive abilities. However, as shown here in section 3.3 as well as in Carpenter et al. (2013), these methods can have heterogeneous effects on subjects with divergent cognitive abilities. Furthermore subtle changes in the experimental design can sometimes affect the efficacy of cognitive load manipulations (Kessler and Meier, 2014). This should be kept in mind when comparing treatment effects from different experiments drawing on these methods, especially when they sample from populations with plausibly diverging distributions of unobserved cognitive abilities e.g. a convenience sample of students and a field sample from the general population.

Even though the interaction is one-shot, I observe higher average cooperation levels for subjects with higher cognitive abilities. Similar findings have also emerged in other studies using finitely repeated (Burks et al., 2009; Jones, 2008) cooperation tasks or non-strategic (Chen et al., 2013) sharing tasks. From a theoretical perspective these findings are surprising: Such one-shot settings lack the *shadow of future interactions*, that can discipline free-riders and hence make cooperation a profitable and rational strategy (Axelrod and Hamilton, 1981). Consequently in several of the existing experiments - including this one - smarter subjects actually appear to be less able to adapt their behavior to the specific incentives created by an one-shot interaction in the lab. Based on the evidence provided here, one can only speculate why this could be the case. Positive

[12] German high-school grades could be a weaker proxy for general abilities than SAT scores, as there is some variation in the difficulty between different federal states

contributions in one-shot PGGs have sometimes been attributed to the possibility that some subjects fail to discriminate between the lab context and their daily interactions and thus behave "as if" the PGG was repeated. If one assumes that this failure is unrelated to cognitive abilities, this could partly resolve the seeming contradiction. Another potential explanation could be, that more intelligent subjects are aware of their higher future earning potential, so that acts of unconditional altruism seem more affordable. A similar argument is made in Millet and Dewitte (2007), who propose that altruism can serve as a costly signal for intelligence.

Obviously, it would be of substantial interest to learn more about the question of why smarter subjects tend to cooperate more. One behavioral channel through which CRT-scores could be related to cooperation is via time preferences and self-control. Subjects with higher cognitive abilities have been shown to be more patient (Frederick, 2005) and to be able to exert higher levels of self-control (de Wit et al., 2007). Experimental evidence as well as theoretical considerations also suggest that both of these attributes could drive behavior in PGGs. Martinsson et al. (2014) demonstrate that subjects with higher self-control cooperate more. Similarly Fehr and Leibbrandt (2011) show how patience assessed by a lab measure predicts cooperative behavior in the field. Thus exploring the relationship between self-control, cognitive abilities and individual discount rates could be of great interest for further research.

One potential limitation of this study is that both through using the CRT and through sampling from a student population, the observed variation of cognitive abilities is fairly limited. Yet taking this into consideration, it is rather even more surprising to which extent subjects behave differently according to their CRT-score.

References

Andreoni, J. (1995). Cooperation in public-goods experiments: kindness or confusion? *The American Economic Review*, 891–904.

Axelrod, R. and W. D. Hamilton (1981). The evolution of cooperation. *Science 211*(4489), 1390–1396.

Bayer, R.-C., E. Renner, and R. Sausgruber (2013). Confusion and learning in the voluntary contributions game. *Experimental Economics* (16), 1–19.

Ben-Ner, A., F. Kong, and L. Putterman (2004). Share and share alike? gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology 25*(5), 581–589.

Benjamin, D. J., S. A. Brown, and J. M. Shapiro (2013). Who is behavioral? cognitive ability and anomalous preferences. *Journal of the European Economic Association 11*(6), 1231–1255.

Burks, S. V., J. P. Carpenter, L. Goette, and A. Rustichini (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences 106*(19), 7745–7750.

Burnham, T. C., D. Cesarini, M. Johannesson, P. Lichtenstein, and B. Wallace (2009). Higher cognitive ability is associated with lower entries in a p-beauty contest. *Journal of Economic Behavior & Organization 72*(1), 171–175.

Carpenter, J., M. Graham, and J. Wolf (2013). Cognitive ability and strategic sophistication. *Games and Economic Behavior 80*, 115–130.

Chen, C.-C., I. Chiu, J. Smith, and T. Yamada (2013). Too smart to be selfish? measures of cognitive ability, social preferences, and consistency. *Journal of Economic Behavior & Organization 90*, 112–122.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–474.

de Wit, H., J. D. Flory, A. Acheson, M. McCloskey, and S. B. Manuck (2007). IQ and nonplanning impulsivity are independently associated with delay discounting in middle-aged adults. *Personality and Individual Differences 42*(1), 111–121.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review 100*(3), 1238–60.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Duffy, S. and J. Smith (2014). Cognitive load in the multi-player prisoner's dilemma game: Are there brains in games? *Journal of Behavioral and Experimental Economics 51*, 47–56.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol. 59*, 255–278.

Fehr, E. and A. Leibbrandt (2011). A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of Public Economics 95*(9-10), 1144–1155.

Ferraro, P. J. and C. A. Vossler (2010). The source and significance of confusion in public goods experiments. *The BE Journal of Economic Analysis & Policy 10*(1), 1935–1682.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives 19*(4), 25–42.

Harrison, G. W., M. I. Lau, and E. E. Rutström (2007). Estimating risk attitudes in denmark: A field experiment. *The Scandinavian Journal of Economics 109*(2), 341–368.

Heckman, J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Heineck, G. and S. Anger (2010). The returns to cognitive abilities and personality traits in germany. *Labour Economics 17*(3), 535–546.

Houser, D. and R. Kurzban (2002). Revisiting kindness and confusion in public goods experiments. *The American Economic Review 92*(4), 1062–1069.

Jones, G. (2008). Are smarter groups more cooperative? evidence from prisoner's dilemma experiments, 1959 to 2003. *Journal of Economic Behavior & Organization 68*(3), 489–497.

Kanazawa, S. and L. Fontaine (2013). Intelligent people defect more in a one-shot prisoners dilemma game. *Journal of Neuroscience, Psychology, and Economics 6*(3), 201.

Kessler, J. B. and S. Meier (2014). Learning from (failed) replications: Cognitive load manipulations and charitable giving. *Journal of Economic Behavior & Organization 102*, 10–13.

Ledyard, J. O. (1994). Public goods: A survey of experimental research. *Public Economics*.

Martinsson, P., K. O. R. Myrseth, and C. Wollbrant (2014). Social dilemmas: When self-control benefits cooperation. *Journal of Economic Psychology*.

Millet, K. and S. Dewitte (2007). Altruistic behavior as a costly signal of general intelligence. *Journal of Research in Personality 41*(2), 316–326.

Nielsen, U. H., J.-R. Tyran, and E. Wengström (2014). Second thoughts on free riding. *Economics Letters 122*(2), 136–139.

Oechssler, J., A. Roider, and P. W. Schmitz (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization 72*(1), 147–152.

Rand, D. G., J. D. Greene, and M. A. Nowak (2012). Spontaneous giving and calculated greed. *Nature 489*(7416), 427–430.

Rand, D. G., A. Peysakhovich, G. T. Kraft-Todd, G. E. Newman, O. Wurzbacher, M. A. Nowak, and J. D. Greene (2014). Social heuristics shape intuitive cooperation. *Nat Commun 5*.

Teyssier, S. (2012). Inequity and risk aversion in sequential public good games. *Public Choice 151*(1-2), 91–119.

Tinghög, G., D. Andersson, C. Bonn, H. Bttiger, C. Josephson, G. Lundgren, D. Vstfjll, M. Kirchler, and M. Johannesson (2013). Intuition and cooperation reconsidered. *Nature 498*(7452), E1–E2.

Toplak, M. E., R. F. West, and K. E. Stanovich (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition 39*(7), 1275–1289.

Toplak, M. E., R. F. West, and K. E. Stanovich (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning 20*(2), 147–168.

Volk, S., C. Thni, and W. Ruigrok (2011). Personality, personal values and cooperation preferences in public goods games: A longitudinal study. *Personality and individual Differences 50*(6), 810–815.

Volk, S., C. Thni, and W. Ruigrok (2012). Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior & Organization 81*(2), 664–676.

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics 6*(3).

The author gratefully acknowledges financial support by the German Ministry for Education and Research under grant OIUV1012. He is furthermore thankful for helpful comments at the ESA Zrich, HSC New York, ZEW Mannheim for their valuable comments.